Published in final edited form as:

J Health Econ. 2008 September ; 27(5): 1182–1195. doi:10.1016/j.jhealeco.2008.05.004.

Using Global Ratings of Health Plans to Improve the Quality of Health Care

Jacob Glazer,

Tel Aviv University Boston University

Thomas G. McGuire,

Harvard Medical School

Zhun Cao, and

Center for Multicultural Mental Health Services Research Cambridge Health Alliance

Alan Zaslavsky

Harvard Medical School

Abstract

Global ratings, such as those based on consumer satisfaction, are a commonly used form of report on the performance of health plans and providers. A simple averaging of the global rating by plan members leads to a problem: it gives a plan greater incentives to improve services used by low-cost members than services used by high-cost members. This paper presents a formal model of consumer formation of global ratings and the incentives these rating convey to plans. We use this model to characterize weights on consumer respondents to correct the incentive problem. We implement our proposed solution using data from the Consumer Assessments of Health Care Providers and Systems (CAHPS) and the Medicare Current Beneficiary Survey (MCBS). Our correction is low-cost, easily implemented on an on-going basis, and insensitive to assumptions about why health plans care about quality ratings.

1 Introduction

When choosing a health plan, consumers have ready access to formal reports about plan coverage, price, indicators of quality of care at the plan, and ratings of consumer satisfaction with services the plan offers. One important component of these reports, consumer satisfaction, is often reported in a summary rating. For example, the widely used Consumer Assessments of Healthcare Providers and Systems (CAHPS®) survey asks:

“Using any number from 0 to 10, where 0 is the worst health plan possible and 10 is the best health plan possible, what number would you use to rate your plan?”

Following Hays et al. (1999), we will refer to ratings derived from questions such as these as “global” ratings. Global ratings of this form are commonly asked and reported in other surveys of plans and hospitals.

Quality reports can improve health care markets in two ways, by giving consumers information so they can choose a plan that best meets their needs and by conveying incentives to plans to

improve their rating to attract more enrollees. In this paper we are concerned with the second function of quality reporting; specifically, the incentives conveyed to a plan to provide high quality by a global rating report. We show that a simple averaging of global ratings of plan members weighting each member equally, the current practice in quality reporting, gives a plan incentives to improve some services more than others. After characterizing the problem, we propose a simple way to fix it, and illustrate how this can be done with recent CAHPS survey data on Medicare beneficiaries. The solution we propose is straightforward and readily implementable on an ongoing basis.

If all patients had the same needs and the rating scale of “worst possible” to “best possible” could be interpreted as equivalent to “no healthcare benefit provided” to “all possible benefit provide”, then a system that reports averaged ratings would convey incentives to provide services to all patients equally. However, when some patients have greater healthcare needs (and potential benefits of care) than others, but all report on the same scale, points on the scale no longer represent the same amount of benefit for different patients. Instead the benefit to less needy patients is overvalued (since a small potential benefit is stretched across the entire scale) while the benefit to needier patients is undervalued (since a large potential benefit is compressed into the same scale). A health plan attempting to maximize its reported ratings then has an incentive to undertreat needier patients and overtreat less needy patients. However, this distortion of the scale can be corrected by weighting reports by patients proportionally to their potential benefit from healthcare; this unequal weighting has the same effect as expanding the scale for the needier patients. We approximate the potential benefit for a patient by the mean spending for a patient with similar health-related characteristics in fee-for-service Medicare (as estimated with data from the Medicare Current Beneficiary Survey) and then apply these weights to determine their impact on comparisons of ratings of health plans.

Section 2 illustrates the incentive problem with a simple numerical example. It also includes a review of the literature on the impact of quality reports and the formation of “global ratings.” Following this review, Section 3 presents a model of global ratings that describes the incentive problem and its solution. Section 4 shows how our correction based on weighting responses by expected costs can be done with CAHPS health plan data for Medicare in 2004. The section also uses Medicare data to quantify the distorted incentives introduced by a global rating. Section 5 summarizes the main point of the paper and identifies some next steps for research.

2 Background

2.1 A Numerical Illustration

The following example illustrates the problem and the nature of the potential solution. A health plan has two enrollees, A and B. Person A uses just service S1 whereas Person B uses both S2 and S3. In a baseline situation depicted in Table 1, the plan spends 5 on each of the three services. Enrollees rate quality of each service they use just in proportion to cost and rate a plan in proportion to the average rating they give to each service they use. So Person A's global rating is 5 as is Person B's. The averaged global rating for the two enrollees is thus also 5, shown in the last column of the table. The health plan considers improving the quality of either S1 or S2. Raising spending on S1 to 10 raises the rating of Person A to 10 yielding an averaged global rating of 7.5. Raising spending on S2 to 10 raises the global rating of Person B only to 7.5 because B also uses S3 and yields an average global rating of only 6.25. (The same calculation would go through for raising quality of S3.) Thus, the plan gets more rating per dollar by spending on S1, the service used by the lower-cost enrollee. Balanced incentives to raise quality across services S1 and S2 would be restored if the averaged global rating weighted Person B's response twice Person A's response. The appropriate relative weight comes from the relative spending at baseline.

2.2 Quality Reports and Health Plan Markets

Research on health plan quality reports has focused on whether reports about quality affect consumers' (or employers') choice of health plan. At least some consumers are prepared to spend time assimilating information in the reports and choose a plan on the basis of the information provided (Farley, Short et al., 2002). Jin and Sorenson (2006) find evidence that consumers appreciate and respond to quality in markets for health plans without quality reporting, and in addition, that formal quality reports from the National Committee on Quality Assurance (NCQA) contain information that also affects plan choice. Papers studying introduction of quality reports to federal employees (Wedig and Tai Seale, 2002), Medicare beneficiaries (Dafny and Dranove, 2004), and private employees (Chernew et al., 2004; Beaulieu, 2002, Scanlon et al., 2002) find that reports affect choice. Employers may also change their offerings in response to quality reports (Chernew et al., 2004). The magnitudes of these effects are small, though. One recent paper studied the health plan market around Minneapolis - St. Paul, Minnesota, and found no measurable effect of quality information and plan switching (Abraham et al., 2006). These papers do not study the importance of different types of information contained in quality reports. One paper suggests that consumers favor global satisfaction over reports of technical quality of care (Schultz et al., 2001).

Quality reports are intended to induce plans (or providers) to improve the quality of care. Employing the invisible hand's "pay-for-performance," quality reports that make demand for a plan more responsive to the plan's choice of quality can lead, in market equilibrium, to a higher overall level of health plan performance. This potential impact of quality reports has not been studied in health plan markets, though it has been investigated in connection with hospital quality reports, with findings that suggest hospitals are alert to the selection-related incentives (Dranove et al., 2003).

A quality report could also motivate a plan or provider through an internal concern for organizational/professional prestige. In other words, even if in their market choices consumers ignored a report, managers of a health plan might seek a high rating for their own professional satisfaction. The arguments we make below about incentive problems/solutions with global ratings go through if global ratings motivate plans for this reason as well as for reasons related to demand response by consumers.

2.3 Consumer Formation of Global Ratings

The relationship between plan features controlled by the plan (e.g., waiting time for an appointment, technical quality of care) and the global rating conferred by members governs the incentives a plan faces to maximize a score on a global rating. If consumer ratings are not sensitive to some dimension of plan performance, the plan gets no reward in terms of ratings for improvement. Cognitive psychologists have studied how individuals formulate responses to various types of questions asked in health-related surveys, such as the mental processes employed to recall numbers of health care visits or the nature of comparisons undertaken for a person to decide if their overall health is "excellent, good, fair or poor" (Fienberg, Loftus and Tanur, 1985). As far as we know, there is no consensus in this literature about how global ratings are formed, but drawing on related studies, we anticipate 1) variation in the cognitive process leading to a rating, 2) anchoring of endpoints based on personal or indirect experience of associates, and 3) salience of more recent, unusual events. These general patterns do not unfortunately offer specific guidance in terms of a theoretical model of how enrollees score plans in a global rating question.

CAHPS and other global ratings have been widely collected for a number of years, and a number of studies have investigated empirically the factors that contribute to higher or lower CAHPS scores. On average, enrollees tend to rate plans very highly (Tai-Seale, 2004). In one

study of elderly and CAHPS, more than half the respondents gave their overall health care (using a CAHPS question) a perfect 10; only about 15% of respondents rated their care less than an 8 (Chang et al., 2006). Beyond this, characteristics of enrollees are associated with scoring. Older enrollees score plans higher, and sicker enrollees score plans lower (Zaslavsky et al, 2001). Racial/ethnic minorities tend to give plans a higher CAHPS global rating score (Lurie et al., 2003; Zaslavsky et al., 2000). It is not easy to interpret these results because it may be hard to separate whether different people tend to score the same events differently, or whether within the same plan, different people have different experiences (Zaslavsky et al., 2001; Zaslavsky and Cleary, 2002). Furthermore, the effect of respondent characteristics on ratings can vary across plans (Zaslavsky, Zaborski and Cleary, 2000).

A number of papers have been interested in how technical measures of the quality of care relate to global consumer ratings. Studies have found either weak (Schneider et al, 2001) or no (Chang et al., 2006) between technical quality and consumer rating. Chang et al (2006) did find that better patient-provider communication was associated with higher global ratings.

3 Incentives Created by Global Rating

Based on the literature on managed care plans, this section develops a model of the incentives for plan quality created by a global rating system. This will allow us to characterize in general terms the incentive problems created when global ratings are simply averaged over all enrollees, and then go on to propose a simple solution to the incentive problem.

3.1 A Managed Care Plan Rationing Services

The model presented here incorporates health plan quality reports into the model of plan behavior in managed care proposed by Frank, Glazer and McGuire (2000). Consider a health plan with N members indexed by i . The health plan offers S services. Let m_{is} denote the amount the plan will spend on providing service s to member i , and let: $m_i = \{m_{i1}, m_{i2}, \dots, m_{iS}\}$. The value of the benefits member i gets from the plan is assumed to be

$$v_i(m_i) = \sum_s v_{is}(m_{is}) \quad (1)$$

where $v_{is}(\cdot)$ is the member's valuation of spending on service s , measured in dollars, and where $v'_{is}=0, v''_{is}=0$. The v_i is the sum of the member's valuations of services offered by the plan.

The plan rations the amount of health care a patient receives. Here, we adopt the shadow-price approach to modelling rationing in managed care.¹ The patient must “need” or benefit from services above a certain threshold in order to qualify for receipt of services. Let q_s be the service-specific shadow price the plan sets determining access to care for service s . Then, a patient with a benefit function for service s of $v_{is}(\cdot)$ will receive a quantity of services, m_{is} determined by:

$$v'_{is}(m_{is}) = q_s \quad (2)$$

Let the amount of spending determined by the equation above be denoted by $m_{is}(q_s)$. Note that (2) is simply a demand function, relating the quantity of services to the (shadow) price in a managed care plan. Let $q = \{q_1, q_2, \dots, q_S\}$ be a vector of shadow prices the plan chooses and $m_i(q) = \{m_{i1}(q_1), m_{i2}(q_2), \dots, m_{iS}(q_S)\}$ be the vector of spending member i receives.

¹For discussion of the shadow price approach to rationing in managed care see Frank, Glazer and McGuire (2000), or Keeler, Carter and Newhouse (1998). The alternative is a quantity rationing approach. See Pauly and Ramsey (1999).

Rationing by (2) implies that within a service, plan funds are allocated efficiently across members. Fully efficient rationing requires further that funds be allocated efficiently across all services. This is achieved by equality of each q_s . The (second-best) efficient q_s can be found by maximizing the sum of member benefits $v_i(m_i)$ subject to a plan budget fixed by capitation payments. We designate the shadow price that solves this constrained maximization as q^0 . We now investigate the incentives a global rating system creates for a plan to choose a vector, q , rationing services its members receive and compare it to the efficient rationing q^0 .

3.2 Member Ratings

Plans choice of shadow prices to ration care will in general differ from the optimal (second-best) shadow prices. In order to characterize plan choices we begin by considering how members rate plans in response to plan rationing. We assume members have a conception of the maximum utility they might attain in a plan and a conception of the minimum utility they might attain. (Indeed, the CAHPS global rating question explicitly asks a respondent to consider a “10” to mean “the best plan possible” and a “0” to mean the “worst plan possible”) A member's rating of a plan is then determined by the actual utility a member receives in relation to the maximum and minimum utility. We regard the maximum utility to be the utility attained when rationing is at its most generous, giving members services up to the point of zero marginal benefit and corresponding to the case when $q_s = 0$, all s . Thus,

$$v_i^{\max} = \sum_s v_{is}(m_{is}(0)) \quad (3)$$

Minimum utility is the utility attained when rationing is at the least generous, which we arbitrarily define to be when $q_s = 1$.² Thus,

$$v_i^{\min} = \sum_s v_{is}(m_{is}(1)) \quad (4)$$

Assume that each individual is asked to rate the plan on the interval between 0 and 1 with the interpretation that the higher is the individual's rating of the plan the more satisfied she is with that plan. Let r_i denote individual i 's rating of the plan and let v_i denote individual i 's utility at that plan. If we assume that r_i is strictly increasing in v_i , and it assigns a rating of 0 to $v_i^{\min}$ and a rating of 1 to $v_i^{\max}$, then any such r_i can be approximated by:

$$r_i = \frac{v_i - v_i^{\min}}{v_i^{\max} - v_i^{\min}} \quad (5)$$

The rating function above captures some of the important elements that seem to affect how people actually form their ratings. Members have a conception of the maximum benefit they might attain in a plan and a conception of the minimum benefit they might attain. A member's rating of a plan is determined by the actual benefit a member receives in relation to the maximum and minimum utility. Hereafter, we'll call $v_i^{\max} - v_i^{\min}$ the “potential benefit” of individual i . A feature of the rating function above, that will be very important later on, is that two individuals with the same (dollar value of) benefits will rate the plan differently if their potential benefit is different.

Thus, for any q with elements between 0 and 1, we can rewrite the member's rating:

$$r_i(q) = \frac{\sum_s v_{is}(m_{is}(q_s)) - v_i^{\min}}{v_i^{\max} - v_i^{\min}} \quad (6)$$

²Values $q = 0$ and $q = 1$ have special properties in the model above but these are not important here. We picked these two values to simplify the presentation. $q_{\min}$ and $q_{\max}$ would have worked just as well.

The overall rating of the plan is then given by

$$R(q) = \frac{1}{\sum_i w_i} \sum_i w_i r_i(q) \quad (7)$$

where w_i is a weight assigned to individual i .

3.3 Plan Rationing to Maximize Rating and Profit

Here we consider how a plan would choose to ration services in the presence of a global rating system, assumed to be the only information to the consumer choosing a plan. The natural starting point for such an analysis would be an expression for the profit gained by the health plan as a function of the global rating. Our problem is simplified by recognizing that the global rating in profit maximization is produced efficiently (a form of cost minimization). Suppose R^* is the global rating that maximizes profits. Since R^* is all that is known about the plan by consumers, enrollment and revenue at the plan (which may depend on risk adjustment) are a function of R^* . Fixing R^* , revenues are not otherwise a function of the vector of q_s the plan chooses to attain this R^* . Thus, profit maximization also implies that the plan would seek to attain the rating R^* at the minimum cost.

An equivalent approach to minimizing cost of attaining a certain R^* is to study the dual problem, maximizing global rating subject to a budget. We study the plan's choice of rationing when it seeks to maximize $R(q)$ subject to a fixed spending on services which we call M :

$$\sum_i \sum_s m_{is}(q_s) = M \quad (8)$$

By the first-order conditions and (8) we get that the plan's choice of rationing q^* to maximize rating is given by the solution to the following system of equations:

$$\left[\frac{\sum_i \frac{w_i m'_{il}(q_1)}{v_i^{\max} - v_i^{\min}}}{\sum_i m'_{il}(q_1)} \right] q_1^* = \left[\frac{\sum_i \frac{w_i m'_{is}(q_s)}{v_i^{\max} - v_i^{\min}}}{\sum_i m'_{is}(q_s)} \right] q_s^* \quad s=2, \dots, S \quad (9)$$

and budget constraint, (8).

Stemming from the literature on risk-adjustment and incentives related to adverse selection, a number of papers have stressed that plans have a motive to ration differentially by service, and either measure these incentives empirically (Frank, Glazer and McGuire, 2000; Ellis and McGuire, 2007), or assess their consequences in terms of patterns of enrollment (Cao and McGuire, 2003) or services (Mello, Stearns and Norton, 2002). If existing risk adjustment systems tend to underpay for the sick and overpay for the healthy, plans have an incentive, similar to the one studied here, to underprovide services used by the sick in relation to those used by the healthy. Keenan et al. (2007) study CAHPS data from Medicare and find consumer ratings of health plans consistent with this hypothesis. Sicker consumers at plans paid by capitation rate the quality of care they receive and the quality of their doctor lower in relation to traditional Medicare than do healthier consumers. Our analysis implies that global rating exacerbates the incentives to undersupply services used by the sicker more costly enrollees.

Note that the incentives associated with the plan's choices of rationing in (9) are, however, distinct from the incentives associated with adverse selection usually studied in the literature on managed care plan rationing. With a single global rating being all consumers know about a plan, all enrollees prefer a higher to a lower ranking. Without further assumptions (that we are not making here), the level of the global ranking itself is not a selection device. The

incentives studied here are those stemming from cost-minimization with respect to production of a global rating.

3.4 Optimal Weighting

Recall that the (budget constrained) socially efficient rationing rule is given by $q_s = q_1$ for $s = 2, \dots, S$ and the budget constraint (8). Equations (9) above will enable us to identify the conditions under which the rating formula induces the plan to provide the socially efficient quality profile.

Let

$$w_i^* = k(v_i^{\max} - v_i^{\min}) \quad (10)$$

for every i and for some $k > 0$. From equations (9) above one can see that If $w_i = w_i^*$ for every i , then $q^* = q^0$. If the weight assigned to each individual, w_i , is set in (the same) proportion to her potential benefit, the rating procedure induces the plan to provide the socially efficient quality. Hereafter, we refer to w_i^* from (10) as the optimal weight.

In order to say something more about the optimal weights, consider the potential benefit $v_i^{\max} - v_i^{\min}$. For each service s and individual i , let $\Delta m_s = m_{is}(0) - m_{is}(1)$, the extra spending on service s individual i enjoys in the most generous rationing compared to the least generous. As a linear approximation, the change in the utility of spending to member i on service s as the shadow price falls from 1 to 0 is the area of a triangle with base Δm_{is} , and height of 1. The potential benefit is the sum of the areas of these triangles, one for each service:

$$v_i^{\max} - v_i^{\min} = \frac{1}{2} \sum_s \Delta m_{is}$$

Making a “demand response” assumption that $\frac{\Delta m_{is}}{m_{is}(0)} = \delta$, for all i and for all s ,³

$$v_i^{\max} - v_i^{\min} = \frac{1}{2} \delta M_i, \quad (11)$$

where $M_i = \sum_s m_{is}(0)$, the total spending of person i with rationing at lowest level level $q_s = 0$ for all s .

Our conception of “potential benefit” in (11) resting behind individuals’ ratings of health plans is shown graphically in Figure 1. The figure depicts marginal benefits for one service, s , for two individuals, 1 and 2. Individual 1 receives fewer benefits from service s and individual 2 receives more. The “worst” conceivable health plan rations with a shadow price of 1 and the best plan rations at zero. The key assumptions we have made about utility are that the benefits from each service can be added to get the total; the reduction in spending from an increase in rationing is proportional to spending at a shadow price of zero, and the slope of the marginal benefit function is constant. These assumptions imply that for these two individuals, the difference in utility (measured in dollars) at the worst and the best plans are the shaded triangular areas. These areas are proportional to spending with the shadow price of zero,⁴ and along with the additivity of benefits assumption allow us to express potential benefit as in (11).

Using this definition of potential benefit, and in particular equations (10) and (11), we get that:

³This is essentially that the arc elasticity of demand is the same across services.

⁴This is true, as we noted above for any two shadow prices corresponding to best and worst.

$$w_i^* = kM_i. \quad (12)$$

for any $k > 0$. Thus, if the rating function assigns to each individual i a weight proportional to her total spending at $q_s = 0$ for all s , the plan will maximize its rating at the socially efficient quality level. Note that with the equality of demand response assumption for all i and s , the optimal weight is proportional to spending for any $q_s = q$, not just $q_s = 0$, $s = 1, \dots, S$.

Another feature of the optimal weight deserves note. Plans choose quality levels, so, to be precise, M_i is the *plan's expectation* about individual i 's spending at $q_s = 0$.⁵ (The optimal weighting equalizes plan incentives to increase the ratings of all members as the plan sees it through its expectations about members' relative spending.) Recognizing that optimal weights are plan's expectations about spending, not actual spending, implies that we should use a predictive model of individual spending rather than actual person-by-person numbers. Our cell-based predictions are described below in Section 4.2.

3.5 Conventional Rating, and Incentives to Distort Quality

In practice, global rating mechanisms simply average all scores. Our analysis can demonstrate the distortion in incentives created by such a mechanism. Assume that $w_i = 1$ for all i , then equation (9) becomes:

$$\left[\frac{\sum_i \frac{m'_{i1}(q_1)}{v_i^{\max} - v_i^{\min}}}{\sum_i m'_{i1}(q_1)} \right] q_1^* = \left[\frac{\sum_i \frac{m'_{is}(q_s)}{v_i^{\max} - v_i^{\min}}}{\sum_i m'_{is}(q_s)} \right] q_s^* \quad s=2, \dots, S \quad (13)$$

Noting that $m'_{is}(q) = \delta m_{is}(0)$ and $v_i^{\max} - v_i^{\min} = \frac{1}{2} \delta M_i$ we can rewrite (13) as

$$\frac{q_1^*}{\bar{m}_1(0)} \sum_i \frac{m_{i1}(0)}{M_i} = \frac{q_s^*}{\bar{m}_s(0)} \sum_i \frac{m_{is}(0)}{M_i} \quad s=2, \dots, S \quad (13')$$

where $\bar{m}_s(0) \equiv \frac{1}{N} \sum_i m_{is}(0)$.

In general, the q^* satisfying (13') and (8) will not be the same as the efficient q^o described above. From (13'), we can see that the factor determining how the q 's will be set in relation to some numeraire, say q_1 , is how m_{is} relates to M_i . This is the correlation across members of spending on service s with total spending. Services with a high positive correlation with total spending will have a low value of the summation term in (13') and will be associated with a high q , i.e., more strict rationing. A positive correlation between m_{is} and M_i derives from m_{is} being relatively large. A service with high levels of spending on average will be more correlated with total spending for the definitional reason that total spending is the sum of its components. Furthermore, a positive correlation can come about because some services are associated with need for health care across a range of illnesses. For example, if depression is a common comorbidity for many chronic (and expensive to treat) illnesses such as heart disease and cancer, mental health services will tend to be positively correlated with total spending.

The incentive to distort quality of services by strict rationing comes about because it is cheaper to "produce" high quality ratings by investing in some services in comparison to others. A service that tends to be used by members who use many services (a high correlation between m_{is} , and M_i) will have any investment in the service diluted by the fact that members using a

⁵For more discussion of the role of plan expectations in characterizing plan incentives see Glazer and McGuire (2002).

lot of care are influenced by other factors in ratings. Conversely, investment in a service used by members who make little use of the plan will tend to yield a higher return.

3.6 Reporting More Than One Global Rating

One might naturally ask how our analysis would change if each plan reported more than one global rating. Before considering this, note that within our model, the optimal global rating achieves the efficient choice of rationing. Unless we change the model in some way, there is nothing to be gained by more than one report. Also, within the model presented so far, we know there is something to be lost. If there were to be a report about each service corresponding to the full-information allocation, incentives for adverse selection would be created and the plan choice of rationing would in general be inefficient. Plausible changes in the model could be made introducing heterogeneity among potential enrollees that would likely alter the conclusion that a single global rating is optimal. In this section, we make some preliminary comments about potential extensions to our analysis recognizing heterogeneity among potential enrollees.

The simplest form of heterogeneity would be if there were two plans that differ according to an observable characteristic (location), and enrollees were heterogeneous in their valuation of this characteristic. If this heterogeneity in taste were distributed independently of service demands, our single global rating system (now of each plan) would likely remain optimal. If, however, consumers' valuation of location is correlated with some service demands, there may be some value in separate reports on location-related and location-unrelated service quality. Some incentives for adverse selection could be introduced (depending on the form of risk adjusted payments) in exchange for a better sorting of consumers between the two plans.

Another direction for extension which is more complicated, but which we think is more interesting, is to suppose that the $v(m)$ function describing consumer valuation of services is composed of two parts: a component related to medical need that the plan sees and rations according to, and a component related to preferences that the plan averages over for all the enrollees in the plan. The purpose of the need/preference distinction would be to capture, for example, different willingness to pay for medical services that high and low-income consumers might have that would tend to be disregarded by providers at managed care plans. Managed care providers, for example, might be willing to spend the same amount of time counseling a high or low-income patient about the risk/benefit tradeoff associated with a surgical procedure. This amount of time would be decided at the plan level in response to the average preferences of enrollees. This average might be "too little" counseling for a high-income patient who would be willing to pay for more but cannot given the plans' rationing systems. Other characteristics of consumers (gender, age or even health status itself) might also condition preferences for health care.

Variation in preferences of this form creates a place for more than one type of health plan in terms of the level of rationing, and potentially premium if the costs of the rationing plans vary as well. An efficient reporting system is called upon to achieve the efficient rationing at each (say, two) plans, and, sort consumers efficiently between the plans. In this case, there may be reason to have more than one global report, perhaps by groups of consumers. The decision about group-specific reports would require evaluation of a tradeoff between preference satisfaction and unwelcome incentives associated with adverse selection. However that is resolved, we believe the weighting we propose here will be part of the efficient reporting policy.

4 Application to CAHPS in Medicare

This section applies the ideas from Section 3 to the largest system for global rating of health plans in the U.S., the CAHPS surveys, used nationally for Medicare managed care plans for

almost ten years (Goldstein, Cleary, Langwell et al, 2001). CAHPS is intended to measure the aspects of care consumers can best report on, and which also correlate with clinical measures of quality (Cleary and Edgman-Levitan, 1997; Schneider, Zaslavsky, Landon et al., 2001). CAHPS has been the basis of many studies in health services investigating the determinants of what consumers value in plans and in how consumer valuation affects plan choice.⁶ CAHPS surveys do not include expenditures, however, and cannot therefore be a source of data on the weights required to modify the averaging for global ratings. We use the Medicare Current Beneficiary Survey (MCBS) to supplement CAHPS with expenditure data. The MCBS is also a national survey collected annually containing publicly available information on a sample of Medicare beneficiaries.

Our primary purpose in this section is to show that our idea contained in (12) for weighting consumer plan ratings can be implemented readily. As a secondary purpose, we apply equation (13') to our data to check for the degree of distortion of rationing incentives introduced by a global rating system that weights each consumer equally.

4.1 Data on Plan Ratings

Data are from the 2004 Medicare CAHPS Survey conducted by CMS. We exclude those who disenrolled during the year, were 64 or younger, enrollees in private fee-for-service plans, and those with missing data on any item used to match with the MCBS. After these exclusions, a maximum of 113,807 respondents in 169 plans remain in the data. In this analysis, we report on two CAHPS global rating items that summarize beneficiary-assessed quality and discriminate well among plans, the rating of the plan and the rating of the care received. Both are reported on a 0–10 scale with 10 being best.

The mean scores of health plans on these items are adjusted for beneficiary characteristics that are not under the direct control of the plan, are associated with ratings and have distributions that differ across plans. The standard Medicare CAHPS adjustment model includes age, education, self-reported general and mental health status, the Medicaid dual eligible flag, flags for proxy assistance with or completion of the form by a proxy respondent, and interactions of health status and age with region (Zaslavsky et al., 2001). Note that this adjustment corrects for predictable differences in plan mean ratings due to the distributions of member characteristics, but does not alter the relative weight given to responses from members with different characteristics.

MCBS data used in this analysis are taken from the 1996, 1997, 1999 and 2000 Cost and Use Files. The MCBS is a continuous, multipurpose survey of a representative national sample of the Medicare population, and therefore can be an ongoing source of expenditure weights for CAHPS. In addition to survey questions, such as self-assessed health status, the MCBS contains information about health plan coverage, and for beneficiaries in traditional fee-for-service (FFS) Medicare, merged claims from Medicare Part A (hospital care) and Part B (Medical Providers). Respondents are surveyed for up to three years and then replaced. We restrict our sample to those beneficiaries in traditional Medicare aged 65 or older, living in the 50 States, alive during the year of survey, without ESRD and enrolled in both Part A and B. Aggregated over all four years there are 27,467 person-year observations in the sample.

Table 2 contains descriptive information about the respondents from the two data sources on the variables we used to match expenditures to survey responses. The age and gender distributions in CAHPS and MCBS are very similar. The MCBS traditional Medicare sample appears to be slightly healthier overall than the CAHPS sample with more respondents

⁶The Agency for Healthcare Quality and Research maintains a website devoted to CAHPS containing information about CAHPS surveys and published and unpublished research material. See <https://www.cahps.ahrq.gov>.

reporting their health as “excellent,” or “very good.” The MCBS also contains fewer smokers. The four age categories, two genders, five general health status categories and two smoker categories yield the 100 cells used in this analysis for purposes of weighting.

4.2 Deriving Expenditure Weights

The optimal weight from (12) is proportional to a member's spending in a plan. Managed care plans generally do not have these data, let alone supply them to Medicare on a regular and timely basis. Instead, we use data from Medicare beneficiaries in traditional FFS Medicare, not in managed care plans, to derive relative weights. FFS Medicare data have the advantage that resource use can be assumed to be driven by beneficiary demand and not affected by plans' decisions about management. It is known that overall spending in Medicare managed care plans is lower than in regular Medicare (Brown, Bergeron and Clement, 1993), that higher cost beneficiaries are more likely to be in regular Medicare (Hellinger, 1995), and that higher costs among beneficiaries in regular Medicare are not uniform across service types (Cao and McGuire, 2003). Though patterns of fee-for-service relative spending differ from those shown in managed care plans, it is not clear what this implies for any bias in relative total spending by cell. The optimal weight need only be proportional to spending, so an overall shrinkage in total spending in managed care plans may not materially alter the desired weights.

The optimal weight is the plan's expected spending for an individual based on their characteristics. We develop and apply a simple cell-based prediction for each person in the CAHPS data. Since we figure spending from one data set, the MCBS, and seek to weight member ratings in another, CAHPS, we can merge estimated spending with CAHPS responses based only on variables common to the two data sets. The common variables we use are age (five categories), gender, self-rated health status (E, VG, G, F, P) and lifetime smoker (Y/N). In total, we have 100 cells and therefore 100 different expenditure weights that we merge with the CAHPS surveys. All of these variables are highly predictive of health care use.

We used the average of the total payment amount, including amount paid by all payers and out-of-pocket payment by the beneficiaries, as a basis for expenditure weights. The overall average spending was \$5,359 for all years together. The average spending in each cell is listed in Appendix A. The mean spending differs quite a bit across the cells (implying substantially different weights) and the cell means move in reasonable patterns. For example, among aged 65–69 year old female non smokers, the averages by self-rated health status are: E:\$1,415, VG:\$2,246, G:\$3,266, F:\$6,917, P:\$10,938. The 100 cell means capture 6.7% of the variance in total expenditure. This compares to the approximately 10% of the variance explained by the Diagnostic Cost Group (DCG) system adopted by Medicare for managed care plan payment that uses demographic, geographic and detailed diagnostic information (van de Ven and Ellis, 2000).

4.3 Results: CAHPS Weighting

Table 3 shows means and standard deviations ratings of care and plan before our weighting, and statistics on the same of ratings after weighting by predicted spending per respondent. Not all respondents answered both questions accounting for the different N's by rating. In the case of both care and plan ratings, the weighted ratings fell slightly, indicating that respondents with higher predicted costs tend to rate plans less highly. In addition, the standard deviation of the plan ratings increased slightly. The original and weighted ratings at the plan level are highly correlated, with Pearson correlation coefficients of .966 and .986 for the care and plan ratings respectively.

We converted the original and weighted ratings into z-scores for each plan to see how much weighting changed plan rankings. The difference in the z-score between the original and

weighted ratings captures the magnitude of the shift due to the weighting. For example, a difference in the z-score of $+0.5$ after the weighting indicates that the plan shifted up $.5$ of a standard deviation of plan means in the ratings due to the weighting. The mean absolute difference in the z-scores for care rating was $.20$ and for plan rating was $.14$. To get a sense of the importance of a $.20$ standard deviation shift, when evaluated at the mean, $.20$ corresponds to an increase of 8.0 percentile rankings (i.e., the mean change shifts a plan above or below 8.0% of the other plans). At one standard deviation above the mean, a $.20$ shift moves a plan 5.0 percentile points. A $.14$ shift, the mean absolute difference in plan rating due to the weighting, corresponds to a 5.6 percentile shift at the mean and a 3.4 percentile shift at one standard deviation above the mean.

The main purpose of this section was to demonstrate that application of optimal weights could be done readily. We had no expectation about how much of a difference the optimal weighting would make in plan rankings. The increase in the spread of scores is probably a good thing since there is such tight clustering at the high end in members' scores in general. The difference in the ranking of plans appears to be noticeable, but slight.

4.4 Results: Incentives in Unweighted Global Ratings

As a secondary empirical analysis, we used the MCBS data, categorized by service, to investigate the incentives for a plan to ration services to maximize an unweighted global rating score. We were interested in whether the shadow prices implied by (13') differed by service; a higher shadow price implying more strict rationing. Differences in shadow prices across services means a plan has incentives created by global ratings to ration some services more strictly than others.

It is not clear how services should be grouped for purposes of studying plan incentives to ration. Conceptually, one would want to group services by the tools available to ration. For example, a plan might decide about the extent and quality of its network of contracting radiologists, making this a category of services subject to incentives. If all non primary care specialists were contracted for together, the unit of analysis should be more aggregated.⁷ Here we choose a high and a low level of aggregation to investigate incentives created by unweighted global rating. Note that the optimal weights can be figured without knowing how "services" should be defined for purposes of rationing.

Our first, more aggregated analysis distinguishes only three services: Inpatient hospital care (Part A), Outpatient hospital care (Part A) and Medical provider (primarily physician) care (Part B). Applying the formula contained in (13'), the shadow prices for these three aggregated services are Inpatient hospital ($.56$), Outpatient hospital ($.27$) and Medical providers ($.30$), indicating that the incentives are to ration inpatient hospital care about twice as strictly as outpatient hospital and medical provider care. This is not surprising, as use of inpatient care is likely to be more highly correlated with total spending.

We next studied services at a disaggregated level. Any Part A or Part B service for which average spending exceeded \$100 per year was treated as a separate category. An "Other Part A" and "Other Part B" category was the sum of the spending in all the categories less than \$100 each. We also separated mental health spending in Part A and Part B because that service has been found to be subject to incentives to underprovide in other analyses (e.g., Cao and McGuire (2003), Ellis and McGuire (2007)). Results are contained in Table 4. Most inpatient hospital shadow prices by more detailed service are around $.5 - .6$. Most medical provider services are around $.2 - .3$. Notably, however, oncology has a very high shadow price of $.87$,

⁷See Glazer and McGuire (2002) for more discussion of the designation of a "service" in the context of optimal risk adjustment analysis.

and the medical specialties with the next highest shadow prices are urology (.37) and cardiology (.36). Higher spending by enrollees who make use of these specialties creates an incentive to ration them more strictly than other physician care.

The incentive to selectively supply services due to global ratings is analogous to adverse selection incentives stemming from capitation payments to plans. If a plan is rewarded by revenue weighted equally for all enrollees (i.e., capitation without risk adjustment), the plan has incentives to provide high quality of care for services used by low-cost enrollees and incentives to provide low quality of care to services used by high-cost enrollees (Frank, Glazer and McGuire, 2000). Analogously, if the plan is rewarded by global ratings weighted equally for all enrollees, the plan has incentives to provide high quality of care for services used by low-cost enrollees, this time because improving care for people who don't use much care is cheaper than improving care for enrollees who use a lot of care. To continue with the analogy, risk adjusting capitation payments can address service-level adverse selection problems, and as we have shown, weighting global ratings by expected expenditures solves the adverse selection problem associated with global ratings.

Other research has characterized incentives created by capitated payments to ration at the service level in Medicare managed care. Mello, Stearns and Norton (2002) find evidence for stricter rationing of Part A than Part B in Medicare. Cao and McGuire (2003) and Ellis and McGuire (2007) use a national 5% sample of beneficiaries and more disaggregated categories. Pulmonary disease, oncological services and inpatient psychiatric care are among the services found by Ellis and McGuire (2007) to be subject to underprovision because of incentives in capitated payments.

5 Conclusions and Directions for Future Research

Global ratings are common in quality reports on plans and providers, and have the advantage that they are easy to collect and understand. As we show here, if global ratings of a plan are a simple average of the ratings of the members, a plan's incentives to improve quality are inadvertently distorted towards services used by low-cost members and against those used by high-cost members. Such a distortion is particularly worrisome because it would reinforce distortions caused by imperfect risk adjustment working against the same set of services (Frank, Glazer and McGuire, 2000). Fortunately, there is an easy intuitive fix for this problem with global ratings. Weighting responses by expected expenditures, giving more say in the global rating to enrollees who cost more, eliminates the distortion and restores balance in incentives to improve quality. This prescription is effective even without knowing what the right way to classify "services" is within a health plan and without knowing whether a plan is motivated by the desire to attract consumers or simply a professional concern with high ratings. Using MCBS weights in CAHPS has a very low cost. The weights are computed in a one-time operation from publicly available data (See Appendix A) and can be used repeatedly to weight responses.

Global or summary ratings have another, related advantage in dealing with adverse selection. Global measures can either be derived from direct questions ("Overall, how would you rate your plan?") or constructed by weighting dimensions of plan or provider performance collected separately into a summary measure (Shahian et al. 2001). A summary or global measure reinforces pooling in health insurance markets. Everyone, no matter what their pattern of service demand, prefers a higher to lower rated plan. This tendency toward pooling is in contrast to rating systems that report on components of plan services (e.g., mental health care, pediatric care, diabetes care) that can contribute to separation according to demand for primary versus some forms of chronic care (for example). Furthermore, by choosing weights to put on elements of plan or provider performance, a regulator can construct a summary measure to counter adverse selection incentives that may arise from other forces, such as competition between

plans in a capitated payment environment (Glazer and McGuire, 2006). Overall, a well-constructed global measure of health plan performance is a promising basis for plan quality reporting.

When global ratings are based on “consumer satisfaction” or some other enrollee-reported subjective measure of performance, interpretation of the rating by either plan administrators or potential enrollees depends on an understanding of how consumers decide to rate a plan high or low. In the course of developing our argument here, we made assumptions about how consumers decide on a rating, the key ones being that it costs more to “move” the rating of a person who uses more services in a plan, and that ratings are related to quality of the services a persons uses. While these assumptions seem reasonable on their face, they are certainly open to question. In any case, study of the quantitative relationship between what a plan does in terms of quality (in many dimensions) and how consumers rate a plan are warranted. This is useful information to plan administrators who must work to improve their plan, consumers who must decide what plan to join, and regulators and payers who must set the terms on which these choices are made.

Acknowledgments

This research was supported by grant R34 MH071242 from the National Institute of Mental Health and the Henry Crown Institute of Business Research (Glazer). Zaslavsky's work was supported by contract number # 500-95-007 from the Centers for Medicare and Medicaid Services (CMS). Cao's work was supported by grant R03 MH71602 from the National Institute for Mental Health. We thank Amy Heller and Elizabeth Goldstein at CMS and the CAHPS project team at Westat for their work that developed the CAHPS datasets used here. We are grateful to Paul Cleary, Richard Frank and Joseph Newhouse for helpful discussion and comments on an earlier draft. Opinions in the paper are the authors' own.

APPENDIX A

APPENDIX A

Crosstab of total expenditure on age, gender, health status, smoking status

agecat	female	genhealth	smoke_lt	Total Expenditure	
				Mean	Std. Err.
age 65–69	male	Excellent	non-smoker	908.11	136.93
age 65–69	male	Excellent	smoker	2341.95	380.47
age 65–69	male	Very Good	non-smoker	2373.50	641.78
age 65–69	male	Very Good	smoker	2759.77	266.90
age 65–69	male	Good	non-smoker	3635.23	622.07
age 65–69	male	Good	smoker	5160.92	543.27
age 65–69	male	Fair	non-smoker	4110.74	1284.10
age 65–69	male	Fair	smoker	8500.75	1137.83
age 65–69	male	Poor	non-smoker	12557.54	4240.11
age 65–69	male	Poor	smoker	16823.99	3640.15
age 65–69	female	Excellent	non-smoker	1415.21	163.49
age 65–69	female	Excellent	smoker	1835.39	390.66
age 65–69	female	Very Good	non-smoker	2246.16	263.35
age 65–69	female	Very Good	smoker	2159.98	196.42
age 65–69	female	Good	non-smoker	3265.90	327.01
age 65–69	female	Good	smoker	4673.40	480.71
age 65–69	female	Fair	non-smoker	6916.72	857.02
age 65–69	female	Fair	smoker	8168.95	754.69
age 65–69	female	Poor	non-smoker	10938.31	2425.01
age 65–69	female	Poor	smoker	14843.89	1574.41
age 70–74	male	Excellent	non-smoker	1498.31	226.30
age 70–74	male	Excellent	smoker	2782.09	315.91
age 70–74	male	Very Good	non-smoker	2669.73	504.27
age 70–74	male	Very Good	smoker	4145.65	292.72
age 70–74	male	Good	non-smoker	4432.63	537.58
age 70–74	male	Good	smoker	5328.99	365.68
age 70–74	male	Fair	non-smoker	8850.01	2305.73
age 70–74	male	Fair	smoker	7580.99	562.08
age 70–74	male	Poor	non-smoker	15624.68	4271.11
age 70–74	male	Poor	smoker	18245.22	2435.61

agecat	female	genhealth	smoke_lt	Total Expenditure	
				Mean	Std. Err.
age 70–74	female	Excellent	non-smoker	2004.73	311.76
age 70–74	female	Excellent	smoker	2673.07	309.49
age 70–74	female	Very Good	non-smoker	2809.16	218.95
age 70–74	female	Very Good	smoker	2860.03	209.69
age 70–74	female	Good	non-smoker	4561.33	347.41
age 70–74	female	Good	smoker	5720.15	514.65
age 70–74	female	Fair	non-smoker	6956.66	528.49
age 70–74	female	Fair	smoker	8282.56	774.92
age 70–74	female	Poor	non-smoker	10896.33	1665.63
age 70–74	female	Poor	smoker	15454.93	1876.04
age 75–79	male	Excellent	non-smoker	2842.41	646.07
age 75–79	male	Excellent	smoker	3166.11	336.91
age 75–79	male	Very Good	non-smoker	4379.71	584.86
age 75–79	male	Very Good	smoker	4137.61	300.21
age 75–79	male	Good	non-smoker	5689.90	603.32
age 75–79	male	Good	smoker	5918.69	412.92
age 75–79	male	Fair	non-smoker	7834.88	1238.13
age 75–79	male	Fair	smoker	9384.35	866.68
age 75–79	male	Poor	non-smoker	11729.17	3594.97
age 75–79	male	Poor	smoker	16246.19	2271.35
age 75–79	female	Excellent	non-smoker	2258.61	289.87
age 75–79	female	Excellent	smoker	2963.89	405.18
age 75–79	female	Very Good	non-smoker	3201.48	184.62
age 75–79	female	Very Good	smoker	4006.53	331.07
age 75–79	female	Good	non-smoker	4872.26	298.09
age 75–79	female	Good	smoker	5141.31	331.07
age 75–79	female	Fair	non-smoker	7914.13	675.25
age 75–79	female	Fair	smoker	7997.21	695.41
age 75–79	female	Poor	non-smoker	11570.12	1345.45
age 75–79	female	Poor	smoker	13807.11	1874.65
age 80–84	male	Excellent	non-smoker	3419.91	543.35
age 80–84	male	Excellent	smoker	4001.81	446.21
age 80–84	male	Very Good	non-smoker	3520.34	413.85
age 80–84	male	Very Good	smoker	5345.29	619.01
age 80–84	male	Good	non-smoker	5950.74	655.50
age 80–84	male	Good	smoker	6726.35	429.46
age 80–84	male	Fair	non-smoker	7957.58	888.35
age 80–84	male	Fair	smoker	8820.60	640.69
age 80–84	male	Poor	non-smoker	16718.25	3117.36
age 80–84	male	Poor	smoker	16678.33	1674.73
age 80–84	female	Excellent	non-smoker	2638.37	333.08
age 80–84	female	Excellent	smoker	3127.23	360.47
age 80–84	female	Very Good	non-smoker	3880.60	278.25
age 80–84	female	Very Good	smoker	4222.59	534.94
age 80–84	female	Good	non-smoker	5140.77	274.33
age 80–84	female	Good	smoker	5909.91	409.80
age 80–84	female	Fair	non-smoker	7248.82	487.26
age 80–84	female	Fair	smoker	6703.35	504.41
age 80–84	female	Poor	non-smoker	10843.34	1083.41
age 80–84	female	Poor	smoker	11855.63	1693.45
age 85+	male	Excellent	non-smoker	3651.95	762.52
age 85+	male	Excellent	smoker	3937.71	536.31
age 85+	male	Very Good	non-smoker	4897.20	624.99
age 85+	male	Very Good	smoker	5378.70	593.23
age 85+	male	Good	non-smoker	4644.57	513.41
age 85+	male	Good	smoker	6618.48	514.69
age 85+	male	Fair	non-smoker	6689.37	877.09
age 85+	male	Fair	smoker	7835.95	682.15
age 85+	male	Poor	non-smoker	9526.57	1925.70
age 85+	male	Poor	smoker	10861.96	1434.27
age 85+	female	Excellent	non-smoker	3312.23	291.47
age 85+	female	Excellent	smoker	3420.80	418.71
age 85+	female	Very Good	non-smoker	3750.47	208.44
age 85+	female	Very Good	smoker	4847.81	548.21
age 85+	female	Good	non-smoker	5443.77	254.87
age 85+	female	Good	smoker	5621.96	554.12
age 85+	female	Fair	non-smoker	6288.66	437.47
age 85+	female	Fair	smoker	8535.99	825.29
age 85+	female	Poor	non-smoker	8899.71	710.13
age 85+	female	Poor	smoker	9955.20	1403.42

Source: Medicare Current Beneficiary Survey (1996–2000).

The statistics are calculated by authors.

References

1. Abraham JM, Feldman R, Carlin C, Christianson J. The effect of quality information on consumer health plan switching: Evidence from the Buyers Health Care Action Group. *Journal of Health Economics* 2006;25:762–781. [PubMed: 16704882]
2. Beaulieu ND. Quality information and consumer health plan choices. *Journal of Health Economics* 2002;21(1):43–63. [PubMed: 11845925]
3. Brown R, Bergeron JW, Clement DG. Do Health Maintenance Organizations Work for Medicare? *Health Care Financing Review* 1993;15:7–24. [PubMed: 10133710]
4. Cao Z, McGuire TG. Service-level Selection by HMOs in Medicare. *Journal of Health Economics* 2003;22:915–931. [PubMed: 14604553]
5. Chang JT, Hays RD, Shekelle PG, MacLean CH, et al. Patients' Global Ratings of Their Health Care are not Associated with the Technical Quality of Their Care. *Annals of Internal Medicine* 2006;144:665–672. [PubMed: 16670136]
6. Chernew ME, Gowrisankaran G, McLaughlin C, Gibson T. Quality and employers choice of health plans. *Journal of Health Economics* 2004;23:471–492. [PubMed: 15120466]
7. Cleary PD, Edgman-Levitan S. Health Care Quality: Incorporating Consumer Perspectives. *Journal of the American Medical Association* 1997;278(19):1608–1612. [PubMed: 9370508]
8. Dafny L, Dranove D. Do Report Cards Tell People Anything They Don't Already Know? The Case of Medicare HMOs. NBER Working Paper. 2004#1142
9. Dranove D, Kessler D, McClellan M, Satterthwaite M. Is More Information Better? The Effects of 'Report Cards' on Health Care Providers. *Journal of Political Economy* 2003;111(3):555–588.
10. Ellis RP, McGuire TG. Predictability and Predictiveness in Health Care Spending. *Journal of Health Economics* 2007;26(1):25–48. [PubMed: 16908083]
11. Farley DO, Short, et al. Effect of CAHPS performance information on health plan choices by Iowa Medicaid beneficiaries. *Medical Care Research and Review* 2002;59(3):319–336. [PubMed: 12205831]
12. Fienberg SE, Loftus EF, Tanur JM. Cognitive Aspects of Health Survey Methodology: An Overview. *The Milbank Memorial Fund Quarterly. Health and Society* 1985;63(3):547–564. [PubMed: 3852078]
13. Frank R, Glazer J, McGuire TG. Measuring Adverse Selection in Managed Health Care. *Journal of Health Economics* 2000;19:829–854. [PubMed: 11186848]
14. Glazer J, McGuire TG. Setting Health Plan Premiums to Ensure Efficient Quality in Health Care: Minimum Variance Optimal Risk Adjustment. *Journal of Public Economics* 2002;84:153–173.
15. Glazer J, McGuire TG. Optimal Quality Reporting in Markets for Health Plans. *Journal of Health Economics* 2006;25(2):295–310. [PubMed: 16352359]
16. Goldstein E, Cleary PD, Langwell KM, et al. Medicare Managed Care CAHPS: A Tool for Performance Improvement. *Health Care Financing Review* 2001;22(3):101–107.
17. Hays RD, Shaul JA, Williams VSL, et al. Psychometric Properties of the CAHPS 1.0 Survey Measures. *Medical Care* 1999;37(3):MS22–MS31. [PubMed: 10098556]
18. Hellinger F. Selection Bias in HMOs and PPOs: A Review of the Evidence. *Inquiry* 1995;32:135–142. [PubMed: 7601511]
19. Jin GZ, Sorensen AT. Information and consumer choice: the value of publicized health plan ratings. *Journal of Health Economics* 2006;25(2):248–275. [PubMed: 16107284]
20. Keeler EB, Newhouse JP, Carter G. A Model of the Impact of Reimbursement Schemes on Health Plan Choice. *Journal of Health Economics* 1998;17(3):297–320. [PubMed: 10180920]
21. Keenan PS, Elliot MN, Cleary PD, Zaslavsky AM, Landon BE. Quality Assessments by Sick and Health Beneficiaries in Traditional Medicare and Medicare Managed Care. 2007unpublished

22. Lurie N, Zhan C, Sangl J, Bierman AS, Sekscenski ES. Variation in Racial and Ethnic Differences in Consumer Assessments of Health Care. *The American Journal of Managed Care* 2003;9(7):502–509. [PubMed: 12866629]
23. Mello MM, Stearns SC, Norton EC. Do Medicare HMOs Still Reduce Health Services Use After Controlling for Selection Bias? *Health Economics* 2002;11:323–340. [PubMed: 12007164]
24. Pauly MV, Ramsey SD. Would You Like Suspenders to Go With That Belt? An Analysis of Optimal Combinations of Cost Sharing and Managed Care. *Journal of Health Economics* 1999;18(4):443–58. [PubMed: 10539616]
25. Scanlon DP, Chernew ME, McLaughlin, Solon G. The impact of health plan report cards on managed care enrollment. *Journal of Health Economics* 2002;21(1):19–41. [PubMed: 11845924]
26. Schneider EC, Zaslavsky AM, Landon B, Lied T, Sheingold S, Cleary PD. National Quality Monitoring of Medicare Health Plans: The Relationship Between Enrollees' Reports and the Quality of Clinical care. *Medical Care* 2001;39(12):1313–1325. [PubMed: 11717573]
27. Schultz J, et al. Do employees use report cards to assess health care provider systems? *Health Services Research* 2001;36(3):509–30. [PubMed: 11482587]
28. Shahian DM, Normand SL, Torchiana DF, Lewis SM, Pastore JO, Kuntz RE, Dreyer PI. Cardiac Surgery Report Cards: Comprehensive Review and Statistical Critique. *Annals of Thoracic Surgery* 2001;76(6):2155–2168. [PubMed: 11789828]
29. Tai-Seale M. Does Consumer Satisfaction Information Matter? Evidence on Member Retention in FEHBP Plans. *Medical Care Research Review* 2004;61:171–186.
30. van de Ven, WPM.; Ellis, RP. Risk Adjustment in Competitive Health Plan Markets. In: Culyer, AJ.; Newhouse, JP., editors. *Handbook of Health Economics*. Elsevier; North Holland: 2000.
31. Wedig GJ, Tai-Seale M. The effect of report cards on consumer choice in the health insurance market. *Journal of Health Economics* 2002;21(6):1031–48. [PubMed: 12475124]
32. Zaslavsky AM, Hochheimer JN, Schneider ED, Cleary PD, et al. Impact of Sociodemographic Case Mix on the HEDIS Measures of Health Plan Quality. *Medical Care* 2000;38(10):981–992. [PubMed: 11021671]
33. Zaslavsky AM, Zaborski LB, Cleary PD. Does the Effect of Respondent Characteristics on Consumer Assessments Vary Across Health Plans? *Medical Care Research and Review* 2000;57(3):379–394. [PubMed: 10981191]
34. Zaslavsky AM, Zaborski LB, Ding L, Shaul JA, et al. Adjusting Performance Measures to Ensure Equitable Plan Comparisons. *Health Care Financing Review* 2001;22(3):109–126.
35. Zaslavsky AM, Cleary PD. Dimensions of Plan Performance for Sick and Healthy Members on the Consumer Assessments of Health Plans Study 2.0 Survey. *Medical Care* 2002;40(10):951–964. [PubMed: 12395028]

Figure 1.
The "Potential Benefit" in Health Plans Among Enrollees with Different Demands

Person A

	S1		Global Rating		S2		S3		Global Rating	
	Cost	Quality			Cost	Quality	Cost	Quality		
Baseline	5	5		5	5	5		5		5
Improve S1	10	10		10	5	5		5		7.5
Improve S2	5	5		5	10	10		5		6.25

TABLE 2**CAHPS and MCBS**

Respondent Characteristics	CAHPS Percent	MCBS Percent
Age		
65–69	20.4	18.3
70–74	28.3	28.0
75–79	23.9	23.7
80–84	16.6	16.3
85+	10.8	13.6
Gender		
Male	40.9	40.3
Female	59.2	59.7
General Health Status (GHS)		
Excellent	7.5	14.9
Very Good	24.4	28.0
Good	38.9	33.0
Fair	24.1	17.9
Poor	5.0	6.3
Smoking Status		
Non Smoker	48.7	42.6
Smoker	51.3	57.4

Sources: 2004 Medicare Managed Care CAHPS Survey, N=113,807.

MCBS 1996, 1997, 1999, 2000, N=27,467.

TABLE 3

Care and Plan Ratings, Medicare CAHPS 2004 Original (Unweighted) and Weighted

	Original		Weighted	
	Mean	Std. Dev.	Mean	Std. Dev.
Care Rating	8.85	0.23	8.75	0.26
Plan Rating	8.43	0.40	8.34	0.43

The N for care rating is 83,260, and for plan rating is 109,656

TABLE 4**Shadow Prices for Selected Inpatient and Outpatient Services**

	Shadow Price
Inpatient Services	
Diseases and Disorders of the Nervous System	0.52
Diseases and Disorders of the Respiratory System	0.50
Diseases and Disorders of the Circulatory System	0.62
Diseases and Disorders of the Digestive System	0.53
Diseases and Disorders of the Musculoskeletal System and Connective Tissue	0.51
Mental Diseases and Disorders, Substance Use	0.52
Factors Influencing Health Status and Other Contacts with Health Services	0.91
All other Inpatient Hospital Care	0.52
Medical Provider Services	
Primary Care Services	0.24
Cardiology	0.36
Ophtho & Optometry	0.22
Orthopedic Surgery	0.31
Psychiatry	0.29
Radiology	0.30
Urology	0.37
Laboratory	0.24
Oncology	0.87
All Other Medical Provider Care	0.31
Outpatient Hospital Service	0.27

Source: Authors' analysis of MCBS data.